Smoothing Techniques of Global Optimization: Distance Scaling Method in Searches for Most Stable Lennard–Jones Atomic Clusters

JAROSŁAW PILLARDY, LUCJAN PIELA

Quantum Chemistry Laboratory, Department of Chemistry, University of Warsaw, 02-093 Warsaw Pasteura 1, Poland

Received 29 April 1997; accepted 24 July 1997

ABSTRACT: Spatial averaging of the potential energy function facilitates the search for the most stable configuration of a molecular system. Recently some global optimization methods of this kind have been designed in the literature that rely on physical phenomena such as diffusion, wave function evolution in quantum mechanics, Smoluchowski dynamics, evolution in temperature of canonical ensembles, etc. In the present article we highlight the fact that all these methods, when applied to the Gaussian distributions of an ensemble, represent special cases of a set of differential equations involving the spatially averaged potential energy. Their structure suggests that the nature's strategy to cope with the global optimization is robust and differs only in the details in particular applications. The strategy consists of going downhill of the averaged potential energy, removing the barriers, and hunting for low energy regions by a selective increasing of the spatial averaging. In this study we explore the deformation of the potential rather than its averaging. The deformation comes from scaling of atomic distances and reduces the barriers even more effectively than the Gaussian averaging. The position and widths of the Gaussian distribution evolve similarly to the Gaussian density annealing (GDA), but we allow elliptical instead of spherical Gaussians as well as branching of the single trajectory of the system into multiple ones. When the temperature reaches 0 K, one has a number of independent Gaussian distributions, each corresponding to a structure and

Correspondence to: L. Piela

Contract/grant sponsor: University of Warsaw; contract/grant number BST S23/23/96

CCC 0192-8651 / 97 / 162040-10

(usually low) energy of the system. The multiple elliptic-Gaussian distance scaling method has been applied to clusters of argon atoms ($N = 5, \ldots, 31$), a system serving usually as a benchmark domain. The method found the global minima for all but three clusters (of very low energy). The procedure is 20 or more times less expensive than the GDA one.
J Comput Chem **18**: 2040–2049, 1997

Keywords: potential energy smoothing; potential energy deformation; potential energy averaging; global optimization; stable structures; atomic clusters

Introduction

To find the global minimum of the conformational energy of even a medium-sized molecular system represents a real challenge. There is a number of force fields in use[1] that offer the potential energy of the system as a function of the interatomic distances. However, all the force fields lead inevitably to an astronomical number of energy minima, each possibly corresponding to a stable structure. At room temperature only a tiny fraction of them is of any interest, the lowest minimum playing usually the most important role in the thermodynamically stable structure. The question is how to find the giobal minimum among myriad others.

The problem of finding the lowest energy structure was intractable before the advent of fast computers. Even now after a decade of efforts[2] there is no theoretical method that is strong enough to cope with it. Two guiding lines of approach seem to be especially hopeful. One is the simutated annealing (SA) method,[3] and the other one is potential energy smoothing.[4] In the SA the temperature is set very high at first, allowing the system to cross even high barriers. Then a slow cooling (using an arbitrary cooling protocol) has the effect of finding lower and lower energy basins with the hope of finding the lowest one(s). A similar effect is obtained in the potential energy smoothing techniques, in which a spatial averaging or deformation of the potential energy is applied. As a result one obtains an effective potential energy hypersurface, which is by far smoother than the original one (for large ranges of the spatial averaging). A dramatically lower number of minima makes it easier to find the global minimum configuration.

The Gaussian density annealing (GDA)[5] method of global optimization made an important link of the diffusion equation method (DEM)[6] to statistical mechanics. In the GDA one solves the reduced Bloch equation satisfied by the normalized coordinate-dependent density distribution ρ for a system in equilibrium with a thermal bath,

$$\frac{\partial \rho}{\partial \beta} = -(V - \langle V \rangle)\rho, \tag{1}$$

where the averaged potential energy $\langle V \rangle(r) = \int V(r')\rho(r' - r)\, dr'$, V denotes the potential energy of the system, $\int \rho\, dr = 1$, and r stands for the set of coordinates. Importantly, the GDA method is more flexible than the DEM by allowing the range of the averaging of the potential energy to be dependent on the position in space. The GDA combines the best features of the SA and the smoothing techniques (cooling with the progress indicator $\beta = 1/kT$ and smoothing by producing the averaged potential $\langle V \rangle$).

In solving this equation one uses the Gaussian Ansatz[5] in which the total equilibrium density distribution ρ is approximated as a product of Gaussian distributions for individual atoms. The latter ones are products of three 1-dimensional Gaussians,

$$\rho(x_i) = (2\pi M_{2,i})^{-(1/2)} \exp\left(-\frac{1}{2M_{2,i}} (x_i - x_{0,i})^2 \right), \tag{2}$$

for $i = 3(n-1) + 1, \ldots, 3(n-1) + 3$. where $x_{3(n-1)+1}$, $x_{3(n-1)+2}$, $x_{3(n-1)+3}$ (vector r_n) are coordinates of atom n; $r_{0,n}$ ($x_{0,3(n-1)+1}$, $x_{0,3(n-1)+2}$, $x_{0,3(n-1)+3}$) is the center of the Gaussian distribution for atom n, for $n = 1, \ldots, N$, and $M_{2,i}$ values are the second moment components of the Gauss-

ian distribution for the system:

$$M_{2,i} = \int (x_i - x_{0,i})^2 \rho \, dr_1 \, dr_2 \cdots dr_N. \quad (3)$$

In the GDA method the $M_{2,i}$ values for i referring to the same atom were chosen to be identical. The reduced Bloch equation together with the Ansatz describes a flow of the approximated (and normalized) equilibrium density as β changes. This leads to the following ''equations of motion'' for the positions and the widths of the Gaussians[7]:

$$\frac{\partial x_{0,i}}{\partial \beta} = -M_{2,i} \frac{\partial \langle V \rangle}{\partial x_{0,i}}$$

$$\frac{\partial M_{2,i}}{\partial \beta} = -M_{2,i}^2 \frac{\partial^2 \langle V \rangle}{\partial x_{0,i}^2}, \quad \text{for } i = 1, \ldots, 3N. \quad (4)$$

To initialize the solving of eq. (4) one starts at the progress indicator $\beta = 0$, an arbitrary $x_{0,i}$, $i = 1, \ldots 3N$, and some large values of $M_{2,i}$ (of the order of hundreds). The results at the progress indicator approaching infinity are expected to be independent of the initial choice of $x_{0,i}$ and $M_{2,i}$, $i = 1, \ldots 3N$. The atomic positions $r_{0,n}$ at β equal to infinity hopefully correspond to the lowest energy structure of the system.

Method

DEM[6] is probably the most familiar smoothing technique. In this method one treats the original potential energy hypersurface as a temperature or concentration distribution and allows for its evolution in time (t) according to the diffusion equation. When the time increases from zero the hypersurface deforms (t is the deformation parameter) and simplifies, exhibiting for sufficiently long times the single minimum that can be easily found. By using the so-called reversing procedure one is able to trace the position of this minimum back to time equals zero (i.e., to its position on the original hypersurface). Very often this position corresponds to the global minimum of the original hypersurface.[8] Important progress was made when Straub and coworkers,[5,7,9,10] Schütte,[11] and Schelstraete and Verschelde[12] related this procedure to the time-dependent Schrödinger equation,[9] canonical ensemble evolution in temperature,[5] Fokker–Planck or Smoluchowski dynamics,[10] smoothing molecular dynamics trajectory in time,[11] and the Gibbs free energy functional.[12] In all these approaches the density distribution (in the case of the imaginary time Schrödinger equation, the ground state wave function) is approximated by a single Gaussian function. These approaches resulted in a number of global optimization methods. A majority of the methods (DEM, imaginary time Schrödinger equation, GDA, and Smoluchowski dynamics) led to the sets of equations for the evolution of the center and the width(s) of the Gaussian density distribution, which have a general form

$$\frac{\partial x_{0,i}}{\partial \alpha} = -A_i \frac{\partial \tilde{V}}{\partial x_{0,i}},$$

$$\frac{\partial D_i}{\partial \alpha} = B_i - C_i \frac{\partial^2 \tilde{V}}{\partial x_{0,i}^2}, \quad \text{for } i = 1, \ldots, 3N, \quad (5)$$

where α is the progress indicator of minimization changing from 0 to ∞ or from ∞ to 0 (depending on the method), A and C are scaling factors for the gradient and curvature of the effective potential energy $\tilde{V}$ (in general they are nondecreasing functions of $M_{2,i}$), and B is a constant; $A, B, C \geq 0$. D stands for the averaging parameter equal to M_2 (the larger the D, the stronger the averaging). In all methods $\tilde{V} \equiv \langle V \rangle(r) = \int V(r')\rho(r' - r) \, dr'$, where ρ is a product of the Gaussian functions of eq. (2). In the DEM $\alpha = t$, t changes from infinity to zero, $A = \frac{3}{4}t$, $D = t$, $B = 1$, and $C = 0$. In the imaginary time Schrödinger equation procedure[9] $\alpha = it$, it changes from infinity to zero, $A = 2M_{2,i}$, $B = (h/2\pi)^2/(2m)$, (where h is the Planck constant and m is the mass of the moving particle), and $C = 2M_{2,i}^2$. In the GDA procedure $\alpha = \beta = 1/kT$, β changes from 0 to infinity, $A = M_{2,i}$, $B = 0$, and $C = M_{2,i}^2$. In the Smoluchowski dynamics method $\alpha = t$, t changes from 0 to ∞, $A = 1/(m\gamma)$ (where m is the particle's mass and $\gamma > 0$ is a coupling constant with a thermal bath ensuring a constant temperature), $B = 2A/\beta$, and $C = 2AM_{2,i}$.

The equations describe a wide range of physical problems. Note that A and B not only vary with the method employed, but also within a particular method because they depend on m, γ, or T. All the global minimization methods that follow from such physical models have been successfully tested on the Lennard–Jones atomic clusters.

This suggests that nature has a general and robust strategy to cope with the multiple minima problem. The strategy varies only in details when

applied to particular physical situations. Let us consider first what would be the most desirable behavior of a global minimization procedure[13]:

1. the original potential should be replaced by an effective potential $\tilde{V}$ that is smoother than V, because a smoother potential has a (dramatically) lower number of minima;
2. the current position should follow the minus gradient of the effective potential, because this is the direction to a minimum of the current basin;
3. the hills of the effective potential should be removed more effectively than the valleys, because the hills are irrelevant when looking for the global minimum; and
4. when trapped in a local minimum or a flat region of the effective potential, the system should have a possibility to go out of there.

Equation (5) would have all these features if $A, B, C > 0$. In all of the methods the effective potential $\tilde{V}$ is a convolution $\langle V \rangle$ of the original one with a Gaussian function of a variable width. (We use the symbol $\tilde{V}$ instead of $\langle V \rangle$, because later on we will generalize the idea of energy deformation by going beyond an averaging.) The convolution means smoothing by cutting out high frequency components of the original potential[6] and therefore the first requirement is satisfied. In all the methods $A, B, C \geq 0$; however, for some of them their values may be equal to zero. The first equation tells us that the position of the Gaussian distribution is heading toward the minimum of a current basin of the effective potential, differing only by the speed of such a motion for a particular method. This satisfies the second requirement for all of the methods. Further, as a consequence of $C > 0$ (all methods except the DEM, for which $C = 0$), when the system is located on a hill ($\partial^2 \tilde{V}/\partial x^2 < 0$), the hill exhibits a tendency to be averaged out by increasing the averaging range; being in a valley ($\partial^2 \tilde{V}/\partial x^2 > 0$) means decreasing of the averaging. This satisfies the third requirement with the intensity characteristic for the method. B is larger than zero for all the methods except the GDA (where $B = 0$), although the particular value of B depends on the method. For the DEM B is a constant related to the progress indicator, for the imaginary time Schrödinger equation method B is related to the kinetic energy and to the Heisenberg principle, and for Smoluchowski dynamics B depends on the temperature of the thermal bath. $B > 0$ means that the deformation or averaging increases when the system is located in a valley with a small curvature. Therefore, its role is to hunt for a lower basin by increasing the deformation parameter or the range of averaging. Once such a lower energy basin is included in the averaging range, $\tilde{V}$ is no longer flat. The created gradient moves [through the first equation in eq. (5)] the center of the distribution toward the minimum of the just discovered basin, thus satisfying the fourth requirement. However, it is desirable to have B tending to zero when the progress indicator α goes to its limiting value, otherwise the deformation may not reach zero in the current minimum at the end of the procedure. Only the Smoluchowski dynamics method exhibits this important feature, if one assumes that $T \to 0$, when $t \to \infty$. Efficiency of a particular global optimization procedure depends certainly on the A, B, C values. Work on the optimal choice of these parameters is under way in our laboratory.

The goal of the third feature is to reduce hills, which do not all carry any information about the global minimum position, and to preserve the minima, which contain this information. However, the Lennard–Jones potential used widely in all force fields has poles (r^{-12}, r is the interatomic distance) that cannot be averaged out and remain in $\langle V \rangle$ at any stage of the procedure. This is related to an important complication, especially clear for the DEM method: the poles of V (very ''hot'' regions) when cooling according to the diffusion equation heat the neighboring cold regions (minima) and change their location considerably, which means a spoiling of information about the location of the global minimum. In practical calculations[5, 8, 12] this problem has been hidden (rather than solved) by approximating the original potential by a sum of Gaussian functions, thus arbitrarily replacing the poles by some final values. The GDA method uses the Gaussian averaging as does the DEM, making it only more pronounced for maxima and less pronounced for minima, while the DEM treats them on equal footing. Thus, in the GDA method there is also a spoiling of information, where the global minimum position is.

When solving eq. (4) one uses $\tilde{V}$ as the averaged potential $\langle V \rangle$, where the range of the averaging is related to M_2. However, according to the above philosophy it might be sufficient to use as $\tilde{V}$ any deformed (not necessarily averaged) potential with the deformation parameter coupled to D of

eq. (5), provided the larger the deformation is, the smoother the deformed potential $\tilde{V}$. One of the simplest deformations is offered by the distance scaling method (DSM).[14] The method proved to be a low-cost alternative for the averaging techniques, when the potential energy is a sum of two-atom contributions $u(r_{ij})$, where r_{ij} is the distance between the ith and the jth atoms. In the present study the potential is deformed (instead of averaging) by a scaling of the interatomic distance in the following way:

$$\tilde{V} = \sum_{i<j}^{N} u\left[r'(r_{ij}, t)\right],$$

$$r'(r, t) = \frac{r + r_0 t}{1 + t}, \qquad (6)$$

where $r = r_0$ is the point for which $r' = r$ and therefore $u(r') = u(r)$ and $\tilde{V}(r') = V(r)$; for any value of the deformation (or scaling) parameters $t \geq 0$. When $t = 0$, then also $r' = r$, $u(r') = u(r)$, and $\tilde{V}(r'(r, t)) = V(r)$ (i.e., no deformation occurs). In the above formulation the DSM method is applicable only to distance-dependent pairwise analytical interaction potentials, in particular in the present application it is used for the Lennard–Jones potential. When r_0 is chosen as the position of the minimum of u (as in the present article) for the Lennard–Jones atoms, then the scaling leaves the position and the depth ε of the minimum of the two-body potential u unchanged. The deformation removes the pole at $r = 0$ for any $t > 0$. The effect of the scaling of the Lennard–Jones potential is its flattening and approaching the value corresponding to the well depth. Any extension of this conclusion to a multidimensional space (even if the interactions are pairwise additive) may be only qualitative nowadays. Note however that there are two main reasons why an energy hypersurface has (unwanted) maxima: the short-range repulsion and the absence of the long-range attraction. Both reasons seem to be weakened by the DSM method.

In the present method we use the elliptic Gaussian atomic distributions of eq. (2) (instead of the spherical ones of the GDA procedure); that is, every dimension i is characterized by a width $M_{2,i}$ of the Gaussian distribution. As a consequence, the ellipses of all different atoms have the same orientation in the global coordinate system. Additionally, we allow for branching of a single atomic Gaussian distribution into two new ones when, in the course of solving eq. (4), the elliptical distribution for an atom becomes highly anisotropic. One may treat these new distributions as satisfying each its own set of the evolution equations, eq. (4); that is, each elliptical Gaussian distribution is treated independently as if it represented the only density distribution. In this context allowing the branching is nothing else but closing a single trajectory and creating two new ones when some conditions are satisfied. Having the possibility of branching, hopefully one does not miss possible bifurcation points. This naturally increases the probability of finding the global minimum in comparison with single trajectory procedures.

A branching attempt was performed in the following way: when the ratio of the maximum $M_{2,i}$ to the minimum $M_{2,j}$ $(i, j = 1, \ldots, 3N)$ exceeds a threshold (assumed to be equal to 2.0 Å^2) and then reaches its maximum, one makes two structures of the system that differ by equal displacements of the atom with maximum M_2 in the opposite directions along the axis for which M_2 is the largest. The magnitude of the displacement is taken arbitrarily, being 20% of the shortest distance of the atom to be displaced to its nearest neighbor. After that the local minimization is performed from the two starting points in order to check whether they correspond to two different basins of the (deformed) hypersurface. The branching is successful if the two minimizations result in different minima and the positions of the new Gaussian distributions are located in the minima found.

To make a coupling between the Gaussian width $M_2 = D$ of eq. (5) and the deformation parameter t of the DSM, eq. (6), we made these two quantities identical. The corresponding procedure will be hereafter called multiple elliptical-Gaussian distance scaling (MEGDS). To estimate the role of the elliptical versus spherical Gaussians as well as the importance of branching, we tried two other auxiliary methods: the MEGDS without branching, called elliptical-Gaussian distance scaling (EGDS), and additionally allowing only spherical Gaussians, spherical-Gaussian distance scaling (SGDS). For comparison we also use the DSM method in its original formulation (i.e., without any coupling to eq. (5); the deformation parameter set to a large value and then application of the reversing procedure) as well as the GDA method.

One might expect two advantages for taking the DSM type deformation rather than the DEM or GDA type of smoothing (besides a disadvantage of losing the link to statistical mechanics). The first is that the DSM not only reduces the maxima of V,

but it also completely removes its poles (the latter restored only at $t = 0$) while keeping the positions of the minima of the two-body potentials unchanged. This hopefully diminishes the spoiling of information where the global minimum is. The second advantage is that in using the DSM one reduces the computation time by a factor of 20 or more in comparison to that of the methods based on the Gaussian approximation of the potential (DEM, GDA). The main reason for this is the important cost of the Gaussian function evaluation when compared to that of the Lennard–Jones formula.

Application to Lennard–Jones Clusters

The above described MEGDS procedure was applied to the Lennard–Jones clusters of the argon atoms ($r_0 = 3.8$ Å, $\varepsilon = 1.68$ kcal/mol). At first the local minimization procedure from 100 starting points was performed in the MEGDS and the EGDS methods at $\beta = 0$ on the maximally deformed hypersurface, resulting in a number of starting minima. (All M_2 values equal to 2.5 Å^2; a larger M_2 produced a too flat hypersurface.) Table I shows how strongly the distance scaling simplifies the original hypersurface. As one can see the number of minima on the undeformed hypersurface grows very rapidly[15] with the number N of the interacting atoms. Also, it is seen that the number of different local minima obtained from 100 randomly generated starting configurations by local minimization initially increases to about 40–50 for $N = 12$ and does not change very much for larger N. The main message from Table I is that the number of minima decreases considerably upon deformation. For any N (when the deformation reaches its maximum), this number, when 100 randomly generated starting points is used, is usually one or two and is only equal to three in a few cases.

TABLE 1.
Reduction of Complexity of the Potential Energy Hypersurface Obtained by Distance Scaling for Clusters of *N* Lennard–Jones Argon Atoms.

N	n_{def}	n	n_{tot}
5	1	2	2
6	1	2	2
7	1	4	4
8	1	6	8
9	2	8	18
10	2	18	57
11	2	27	145
12	2	40	366
13	2	41	988
14	2	41	3258
15	1	43	10,700
16	3	41	37,053
17	2	46	136,094
18	3	47	529,294
19	1	53	2,180,000
20	2	53	9,504,729
21	2	57	43,885,740
22	3	52	214,560,005
23	3	53	1,110,748,556
24	2	47	6,088,697,375
25	2	46	35,340,660,729
26	2	47	2.17203E + 11
27	2	49	1.41351E + 12
28	3	44	9.74037E + 12
29	2	46	7.10710E + 13
30	1	45	5.49099E + 14
31	2	50	4.49211E + 15

Independent starts of local minimization were performed from 100 randomly generated configurations. The second and the third columns contain the numbers of minima found by local minimization from the obtained starting points on the highly deformed (n_{def}) and the original (n) hypersurfaces, respectively. The scaling parameter $M_2 = 2.5$ Å^2. The fourth column displays the total number of minima (n_{tot}) estimated by an approximate formula.[15]

Table II gives the lowest energies found for $N = 5, \ldots, 31$ (at $\beta = \infty$) by using the MEGDS, EGDS, SGDS, DSM, and GDA and compares the lowest ones known in the literature as well as the results of the local minimization. In all these calculations the previously generated starting points were used. It is seen that the local minimization is remarkably successful despite a huge number of local minima. The reason may be that we started the local minimization from a large dimension of clusters.[16] This temporary constraint supports only a relatively small number of basins, which may be found by the local minimization procedure and evolve to the global minima for $N < 17$. As one may expect, the local minimization has to fail for large N, which is what indeed happens for $N > 16$. Table II shows also that the MEGDS procedure is most successful in finding the lowest energy structures. The method produces the global minimum structures for all N except $N = 8$, 27, and 28, where other low energy minima have been found.

$N = 8$ is known in the literature as being one particularly difficult to obtain, because of a quite peculiar quasidegeneracy of the two lowest minima.[8, 10, 17] Our experience shows that the deterministic methods (like the DEM or DSM) always lead to a higher energy minimum (HM), corresponding to a wide basin failing to find the global one (GM) with a narrow basin. In accordance with this picture, methods applying stochastic perturbations give both minima; however, with a much larger probability to find an HM. It may be that when solving eq. (5) both minima do not emerge by a splitting of a common basin at a particular point in the multidimensional space, but rather that the GM basin emerges at the slope of the much wider basin of the HM. This might be a reason why the MEGDS method finds the HM and fails to find the GM.

As shown in Table II, the results of the auxiliary methods, EGDS and SGDS, as well as those of the DSM and GDA are nearly equal in quality and are substantially worse than those obtained by the MEGDS procedure. Allowing the ellipticity of the Gaussians (EGDS vs. SGDS) increases the numerical stability of the minimization to such an extent that it speeds up the minimization procedure by a factor of 20.

Figures 1 and 2 show how the branching of the Gaussian density occur when β increases. As an example, $N = 11$ is taken. From Fig. 1a it is seen that $\tilde{V}$ increases with β (the opposite happens in the GDA[5]). This is a consequence of the fact that a large scaling removes interatomic repulsions, which are then restored gradually when the temperature (and the deformation parameter) decreases. Fig. 1a shows also that the first branching

TABLE 2.
Comparison of Performance of MEGDS, EGDS, SGDS, DSM, and GDA Methods in Global Minimization of Potential Energy of Clusters of *N* Lennard–Jones Argon Atoms.

N	Global	MEGDS	EGDS	SGDS	DSM	GDA	Minimization
5	**−15.295**	**−15.295**	**−15.295**	**−15.295**	**−15.295**	**−15.294**	**−15.295**
6	**−21.356**	**−21.356**	**−21.356**	**−21.356**	**−21.356**	**−21.356**	**−21.356**
7	**−27.729**	**−27.729**	**−27.729**	**−27.729**	**−27.729**	**−27.729**	**−27.729**
8	**−33.300**	*−33.206*	*−33.206*	**−33.299**	*−33.206*	**−33.300**	**−33.300**
9	**40.510**	**−40.510**	**−40.510**	**40.510**	**−40.510**	**−40.510**	**−40.511**
10	**−47.750**	**−47.750**	**−47.750**	**−47.750**	**−47.750**	**−47.750**	**−47.750**
11	**−55.047**	**−55.047**	**−55.047**	**−55.047**	**−55.047**	**−55.047**	**−55.047**
12	**−63.786**	**−63.786**	**−63.786**	**−63.786**	**−63.786**	**−63.786**	**−63.786**
13	**−74.469**	**−74.469**	**−74.469**	**−74.469**	**−74.469**	**−74.469**	**−74.469**
14	**−80.380**	**−80.380**	**−80.380**	**−80.380**	**−80.380**	**−80.380**	**−80.380**
15	**−87.902**	**−87.902**	*−84.897*	**−87.902**	**−87.902**	**−87.902**	**−87.902**
16	**−95.450**	**−95.451**	*−93.984*	**−95.450**	*−93.984*	**−95.450**	**−95.451**
17	**−103.014**	**−103.014**	**−103.014**	*−102.996*	*−102.996*	*−102.866*	*−102.996*
18	**−111.772**	**−111.772**	**−111.772**	*−111.358*	*−111.358*	*−111.069*	*−111.358*
19	**−122.068**	**−122.068**	**−122.068**	*−119.419*	*−119.419*	**−122.068**	*−118.975*
20	**−129.657**	**−129.657**	**−129.657**	**−129.657**	**−129.657**	**−129.657**	**−129.657**
21	**−137.230**	**−137.230**	*−137.175*	*−137.177*	*−137.175*	*−137.175*	*−137.175*
22	**−145.840**	**−145.840**	**−145.840**	**−145.840**	**−145.840**	*−144.849*	**−145.840**
23	**−155.979**	**−155.979**	**−155.979**	**−155.979**	**−155.979**	**−155.979**	*−153.498*
24	**−163.546**	**−163.546**	**−163.546**	*−163.486*	*−163.486*	*−163.486*	*−162.059*
25	**−171.986**	**−171.986**	*−168.263*	**−171.986**	*−170.984*	*−170.983*	*−170.983*
26	**−181.970**	**−181.970**	**−181.970**	**−181.970**	**−181.970**	**−181.970**	*−179.954*
27	**−189.628**	*−189.547*	*−186.147*	*−189.547*	*−189.547*	*−188.701*	**−189.628**
28	**−197.942**	*−197.867*	*−195.050*	*−197.028*	*−196.973*	*−193.344*	*−195.645*
29	**−207.627**	**−207.627**	*−201.943*	*−201.943*	**−207.627**	*−203.243*	*−205.151*
30	**−215.521**	**−215.521**	*−206.245*	**−215.521**	*−213.250*	*−214.755*	*−213.250*
31	**−224.425**	**−224.425**	*−219.729*	*−223.034*	*−219.098*	*−220.133*	*−223.034*

The second column contains the lowest energy of the clusters (in kcal / mol) found in the literature. The lowest energy in the row is highlighted by the bold face throughout the table; other values are in italic. The third–eighth columns display the results of the MEGDS, EGDS, SGDS, DSM, and GDA methods and of the local minimization.

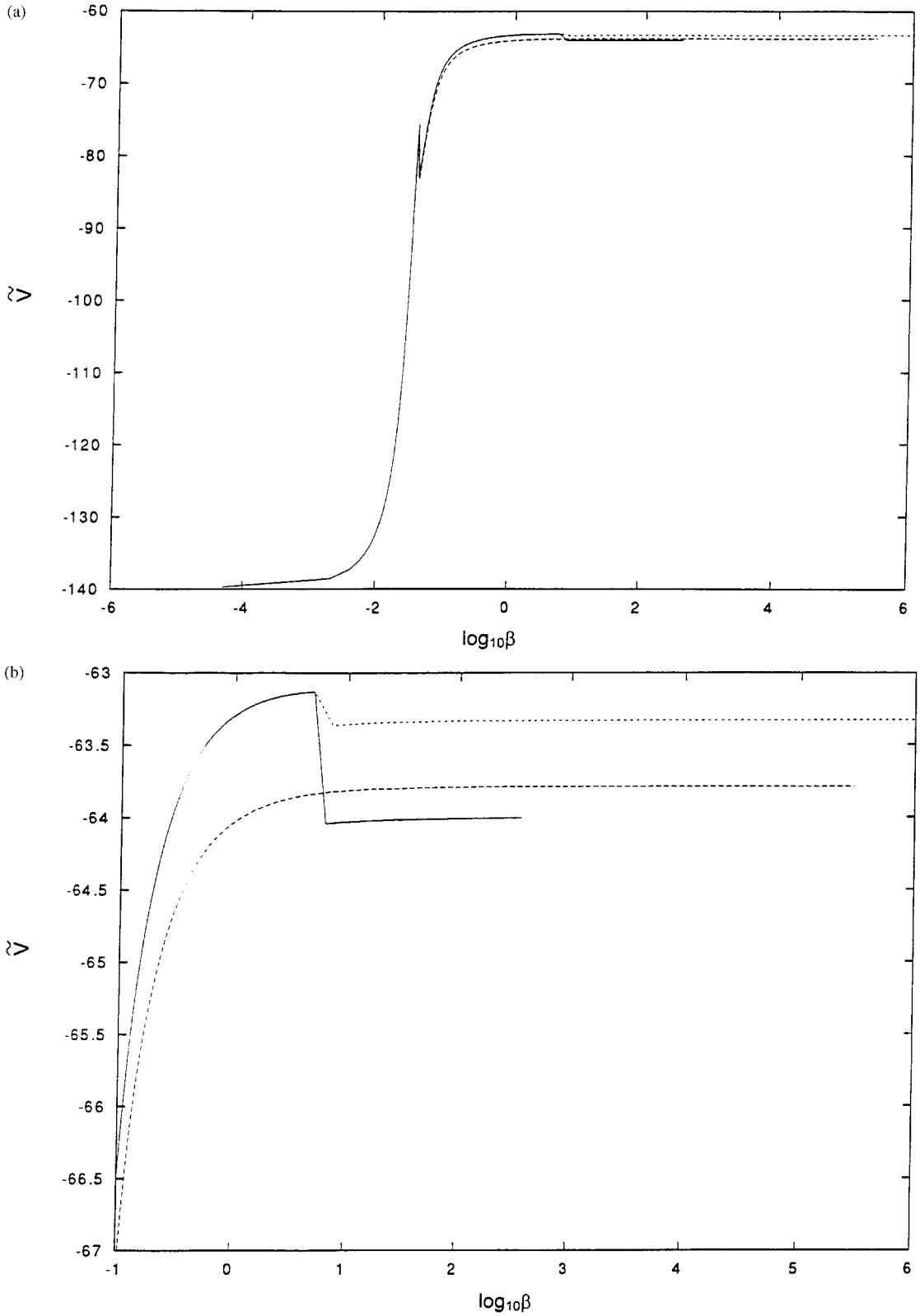

FIGURE 1. (a) The deformed potential energy (in kcals/mol) of Ar_{11} for various trajectories as a function of $\log_{10}\beta$, $= 1/kT$. (b) The details of the plot for larger values of $\log_{10}\beta$ are shown.

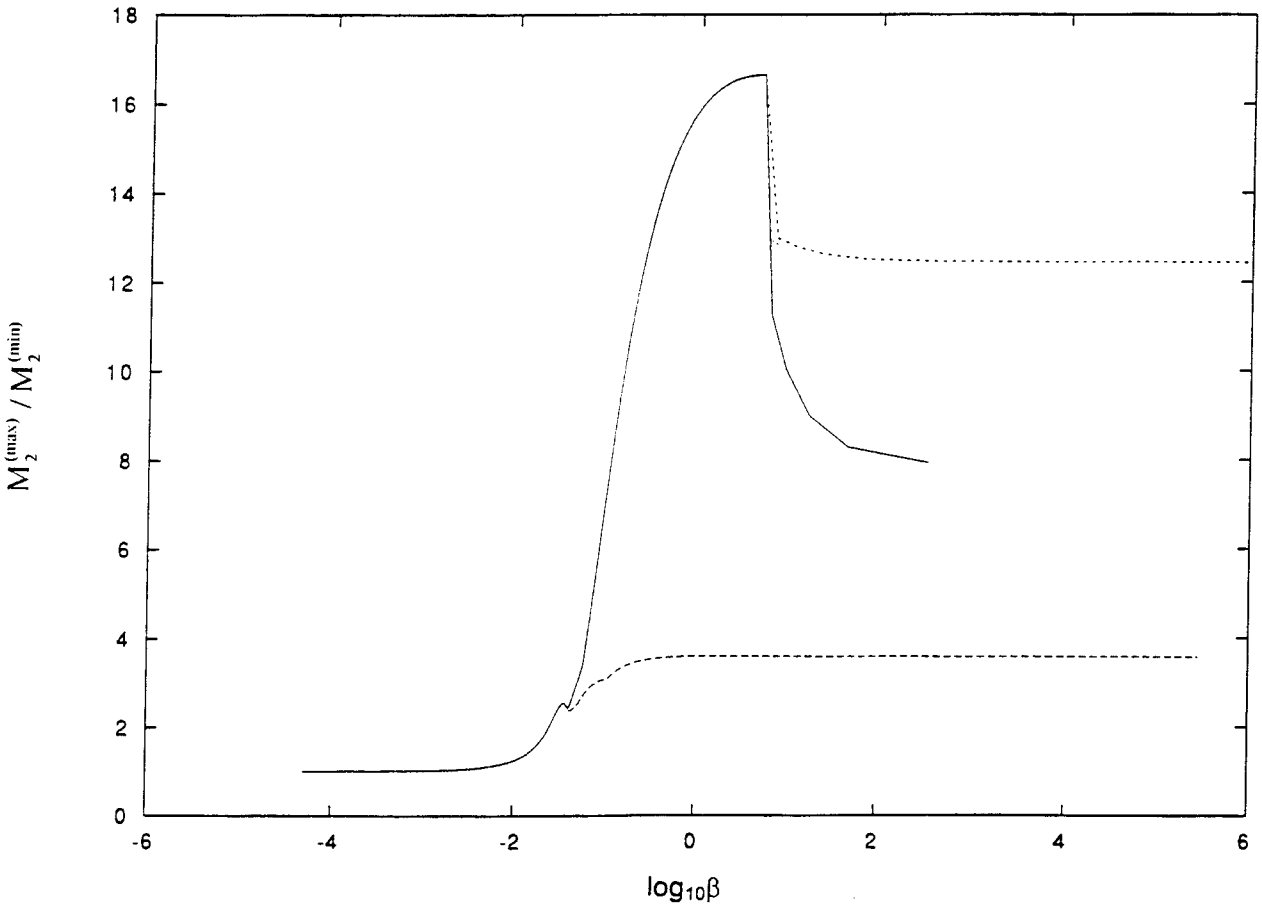

FIGURE 2. Anisotropy of the Gaussian distribution ($M_{2,i\max}/M_{2,i\min}$) for various trajectories during the global minimization; $\beta = 1/kT$.

occurs at $\log_{10}\beta = -1.43$ and is accompanied by a substantial energy lowering of both structures. The higher-energy structure undergoes another branching at $\log_{10}\beta = 0.69$; its details are magnified in Fig. 1b. Interestingly, one of the resulting structures lowers its $\bar{V}$ below that of all other structures and finally ends up as the GM one.

Fig. 2 shows the dependence of the ratio of the maximum to minimum value among all $M_{2,i}$ values. A low value of the ratio means a structure with rather isotropic Gaussian distributions of the positions of individual atoms; a large value indicates a highly anisotropic distribution. In the last case a branching test is performed. Figure 2 shows the extent of the anisotropy and the branching at $\log_{10}\beta = -1.43$ and 0.69. The procedure stops for a given distribution if its maximum width becomes lower than 10^{-8} Å^2.

Conclusions

The conclusions of the present study may be summarized as follows:

1. The global minimization methods stemmed from the diffusion equation. The time-dependent Schrödinger equation in quantum mechanics, canonical ensemble annealing (Bloch equation), and the dynamics of an ensemble coupled to a thermal bath (Smoluchowski dynamics) may be viewed as having the same mathematical structure. The structure reveals the main features of nature's strategy to find the GM: averaging the potential energy, going downhill, averaging out the hills, and

hunting for new lower energy regions of the averaged potential energy.

2. Relying on these features, a method of global minimization was constructed that reduces the hills of the potential energy hypersurface even more effectively than averaging. The method involves distance scaling and allows an elliptical Gaussian distribution for each atom of the system as well as branching of the trajectory of the Gaussian distribution.
3. The method was applied to the argon atom clusters up to Ar_{31} and found the global minimum for all but three cases.

Acknowledgments

This work was supported by Grant BST 523/23/96 from the Department of Chemistry and a grant from the Interdisciplinary Center for Computational and Mathematical Modeling, University of Warsaw.

References

1. A comparison of three of the force fields CHARMM, AMBER, and ECEPP is done for peptides by I. K. Roterman, H. M. Lambert, K. D. Gibson, and H. A. Scheraga, *J. Biomol. Structure Dynam.*, **7**, 421, (1989).
2. See review articles: (a) J. E. Straub, In *New Developments in Theoretical Studies of Proteins*, R. Elber, Ed., World Scientific, Singapore, **1995**; (b) L. Piela, K. A. Olszewski, and J. Pillardy, *J. Mol. Struct.*, **308**, 229 (1994); (c) M. Vásquez, G. Nemethy, and H. A. Scheraga, *Chem. Rev.*, **94**, 2183 (1994); (d) R. Elber, *Curr. Opin. Struct. Biol.*, **3**, 260 (1993).
3. S. Kirkpatrick, Jr., C. D. Gelatt, and M. P. Vecchi, *Science*, **220**, 671 (1983).
4. The first article that discusses a smoothing of the potential energy function by the convolution transformation was F. H. Stillinger, *Phys. Rev. B*, **32**, 3134 (1985). The article deals with the glass transition for amorphous substances and gives general formulas concerning the exponential decrease of the number of minima when the Gaussian smoothing of the potential energy hypersurface is performed. In L. Piela, J. Kostrowicki, and H. A. Scheraga, *J. Phys. Chem.*, **93**, 3339 (1989) a similar idea has been introduced independently with the following differences: smoothing is introduced as going beyond any convolution; a global minimization strategy has been constructed that consists of two steps, convexation of the potential yielding the single minimum, and the so-called reversing procedure that hopefully links the convex potential to the global minimum; and the strategy has been applied numerically.
5. J. Ma and J. E. Straub, *J. Chem. Phys.*, **101**, 533 (1994).
6. (a) L. Piela, J. Kostrowicki, and H. A. Scheraga, *J. Phys. Chem.*, **93**, 3339–3346 (1989).
   (b) J. Kostrowicki and L. Piela, *J. Optimization Theory Appl.*, **69**, 269 (1991).
7. J. Ma, D. Hsu, and J. E. Straub, *J. Chem. Phys.*, **99**, 4024 (1993).
8. J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga, *J. Phys. Chem.*, **95**, 4113 (1991).
9. P. Amara, D. Hsu, and J. E. Straub, *J. Phys. Chem.*, **97**, 6715 (1993).
10. J. E. Straub, J. Ma, and P. Amara, *J. Chem. Phys.*, **103**, 1574 (1995).
11. Ch. Schütte, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Preprint SC 1995, 95-14 (1995).
12. S. Schelstraete and H. Verschelde, *J. Phys. Chem.*, **101**, 315 (1997).
13. One may think of even a simpler feature of an ideal smoothing method. Such a method should produce a coarse grained hypersurface with the single minimum *exactly in the position of the global minimum* of the original hypersurface. Going beyond the convolution idea one may imagine such a method as a ''wheelbarrow smoothing,'' which in one dimension may be described as follows: one creates a trajectory (it will represent a coarse grained function) of the axis of a wheel rolling over the function from *below*. The method has a most precious feature that at any radius of the wheel the coarse grained function has the global minimum exactly at the position of the global minimum of the original function. When the radius is zero one has the original function; when the radius goes to infinity one ends up with a coarse grained function with only a single minimum (when the global minimum is nondegenerate). Work on this problem is in progress in our laboratory.
14. J. Pillardy and L. Piela, *J. Phys. Chem.*, **99**, 11805 (1995).
15. The number of minima was computed from a formula obtained in M. R. Hoare, In *Advances in Chemistry and Physics*, I. Prigogine and S. A. Rice, Eds., Wiley, New York, 1979, p. 40. This was done by fitting to values of $N < 13$ and extrapolating to large N.
16. J. Pillardy and L. Piela, manuscript in preparation.
17. J. Pillardy, K. A. Olszewski, and L. Piela, *J. Phys. Chem.*, **96**, 4337 (1992).